Patterns of Cognitive Decline in Alzheimer's Dementia: An Application of Hierarchical Bayesian Multiphase Models

By

Zachary Denver Langford

Submitted to the Department of Psychology and the
Faculty of the Graduate School of the University of Kansas
in partial fulfillment of the requirements for the degree of
Master's of Arts

David K. Johnson, Ph.D., Chairperson

Committee members

Todd Little, Ph.D.

William Skorupski, Ph.D.

Date defended: December 4, 2012

The Thesis Committee for Zachary Denver Langford certifies
that this is the approved version of the following thesis :

Patterns of Cognitive Decline in Alzheimer's Dementia: An Application of Hierarchical Bayesian Multiphase Models

David K. Johnson, Ph.D., Chairperson

Date approved: December 4, 2012

Abstract

As the Alzheimer's disease process progresses in time measurements of cognitive functioning exhibit nonlinearity. Multiphase models were used to quantify this nonlinearity for thirty-six well characterized individuals($\sim$12 observations per individual over $\sim$15 years in the study) by partitioning each into a healthy aging phase and a diseased phase. This enabled us to detail both the magnitude *and* timing that Alzheimer's disease alters different aspects of cognitive function. Estimation of these models was done using Bayesian methods. Eight different outcomes representing three areas of memory functioning(visual, verbal, working) were used to define a pattern of cognitive decline. The earliest phase change was found to be visual memory($\sim$6 years before diagnosis) and was followed by changes in verbal and working memory beginning roughly four years later.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Disease Processes

Many disease processes have biomarkers that develop over time in a nonlinear fashion - in healthy and diseased phases. T4 counts show a rapid decline close to development of AIDS [Kiuchi et al., 1995], prostate specific antigen spikes before development of prostate cancer [Slate & Turnbull, 2000], and adenosine deaminase exhibits a rapid change in chronic myelogenous leukaemia [Klein JP, 1984]. Alzheimer's disease (AD) is not an exception. In **Figure 1.1** an individual's trajectory of measurements on a visuospatial memory test (top) and a working memory test (bottom) are plotted. For this individual the time that they were clinically healthy is marked

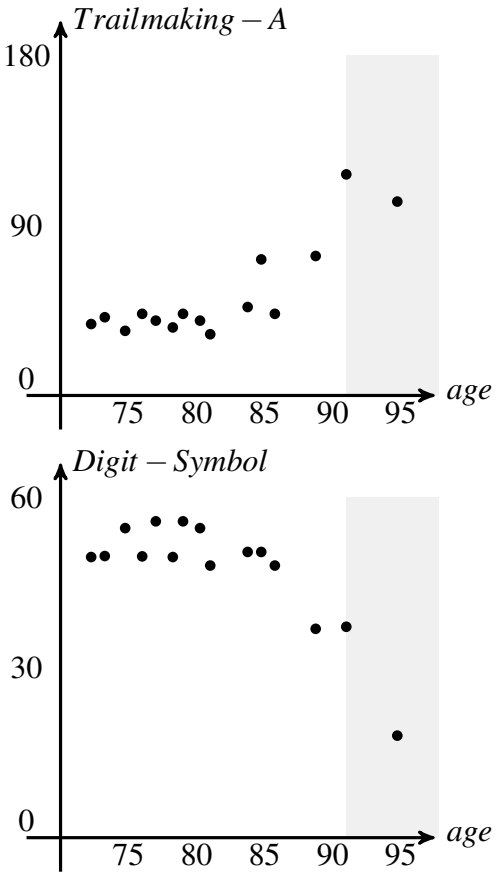

Figure 1.1: Individual Scatterplot

as the white portion and the time that they spent diagnosed with AD is the grey portion. Confirmation of the clinical AD diagnosis was post-mortem. As shown in the figure, this person undergoes a change in both visuospatial and working memory abilities at some points *prior* to clinical diagnosis.

1.2 Purpose

As the biological process of AD progresses in time different cognitive abilities decline. This thesis focuses on the determination of how these disparate processes change over time in AD to define a pattern of cognitive decline. Multiphase models are a class of models that can be used to quantify both the magnitude in which AD alters different cognitive faculties *and* the timing at which this alteration begins as the disease process evolves. In this thesis the cognitive processes in AD are modeled with two phases - one phase demarcating healthy cognition and the other diseased cognition. The changepoint, the parameter of most interest, is the time of the switch between these two phases. Hierarchical Bayesian techniques are used to model both individual development and to relate the individual to the AD population.

1.3 Organization

In section 1.4 an introduction to this methodology will characterize these phase parameters for the visuospatial abilities of the individual in **Figure 1.1**. Chapter 3 will explain how to extend this individual model to a hierarchical model to characterize the AD population. Chapter 2 borrows methods from misclassified multistate models to define a stringent procedure for identifying a *non-diluted* AD population - addressing a weakness in prior research. Both chapters 4 and 5 will discuss the parameter estimates and delineate the pattern of cognitive decline in AD.

1.4 Primer: Modeling a Single Individual

1.4.1 A Model with Linear Phases

Although the full strength of this thesis is seen in the hierarchical models developed in Chapter 3 it is informative to describe the model as fit to a single individual. To do this, we will use the data from the single individuals measurements on the visual memory test shown in **Figure 1.1**. Assume that the individuals jth measurement Y_j is distributed normally with expectation μ_j and precision τ[1], $Y_j \sim N(\mu_j, \tau)$. A multiphase model that is continuous at the changepoint and with two linear phases surrounding the changepoint is given by

$$\mu_j = \begin{cases} \alpha_0 + \alpha_1 t_j & t_j < \delta \\ \alpha_0 + \alpha_1 \delta + \alpha_2 (t_j - \delta) & t_j \geq \delta \end{cases} \tag{1.1}$$

The parameters in the model are shown in **Figure 1.2.** α_0 is the level of Y at the beginning of the first phase. α_1 is the slope of the first linear phase and α_2 is the slope of the second linear phase. δ, the changepoint, is the time of the switch between α_1 and α_2. In this individual analysis the time scale was age. Note that the intercept of the second phase(call it α_3) is a redundant parameter. If the segments are continuous at δ then $\alpha_0 + \alpha_1\delta = \alpha_3 + \alpha_2\delta$ and α_3 can be determined from the others, i.e. $\alpha_3 = \alpha_0 + \alpha_1\delta - \alpha_2\delta$.

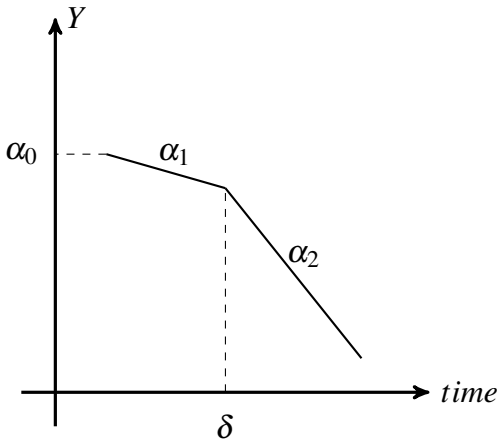

Figure 1.2: Multiphase Model

[1] $\tau = \frac{1}{\sigma^2}$

1.4.2 Bayesian Estimation

The models developed later in this thesis become more highly parameterized than in Equation 1.1, but the nature of parameter estimation does not change. The scheme is explained generally in this section. Let θ denote our parameter vector. Bayesian estimation is concerned with modifying prior knowledge $\pi(\theta)$ of the parameters through the likelihood of the observed data $f(y|\theta)$ to characterize posterior densities $\pi(\theta|y)$

$$\pi(\theta|y) = \frac{f(y|\theta)\pi(\theta)}{\int f(y|\theta)\pi(\theta)d\theta} \tag{1.2}$$

For our simple example, the likelihood $f(y|\theta)$ can be derived from equation 1.1. The elements of the prior vector $\pi(\theta)$ were assigned diffuse conjugate distributions - specifics are not needed at this juncture. That leaves the normalization constant $\int f(y|\theta)\pi(\theta)d\theta$; an integration over all possible values of θ that guarantees the posterior is a probability distribution (and complicates estimation).

Obtaining any specific marginal posterior θ_j from the joint density $\pi(\theta|y)$ involves integrating out all other 'nuisance' parameters

$$P(\theta_j|y) = \int P(\theta_1...\theta_{j-1}, \theta_{j+1}...\theta_k|y)d\theta_1...d\theta_{j-1}, d\theta_{j+1}...d\theta_k \tag{1.3}$$

The most straightforward approach would be to calculate $P(\theta_j|y)$ analytically or numerically. This is intractable in the case of the high-dimensional integrals in this thesis. Gibbs sampling [Geman & Geman, 1984] provides an alternative to direct computation. A Gibbs sampler allows us to set up a Markov Chain Monte Carlo (MCMC) algorithm to generate a sample $\Theta_{j1}, ..., \Theta_{jm} \sim P(\theta_j|y)$ without requiring $P(\theta_j|y)$ - given that we can state the complete conditional densities $P(\theta_j|y, \theta_1...\theta_{j-1}, \theta_{j+1}...\theta_k)$. That is, we can obtain a posterior density to a desired accuracy (determined by the length of m) if we can specify all the conditional densities. Implementation of Gibbs sampling using directed acyclic graphs in JAGS [Plummer & Plummer, 2003] will be discussed in Chapter 3.

1.4.3 Parameter Estimates

The individuals estimated multiphase structure has been plotted in **Figure 1.3**. The mean estimates and 95% credible intervals of the posterior distributions of interest are as follows; $\alpha_0 = 38.37(31.82, 44.03)$, $\alpha_1 = -0.26(-1.17, 1.16)$, $\alpha_2 = 5.28(4.01, 6.25)$, $\delta = 81.68(79.75, 83.67)$. A credible interval is simply an area where some percentage of the posterior simulation is contained. Looking at the graphic we see that for this individual their performance on the Trailmaking-A test stays roughly constant until they hit the estimated changepoint at 81.68 years of age and then their performance quickly becomes worse - approximately 8 years before clinical diagnosis (the grey region).

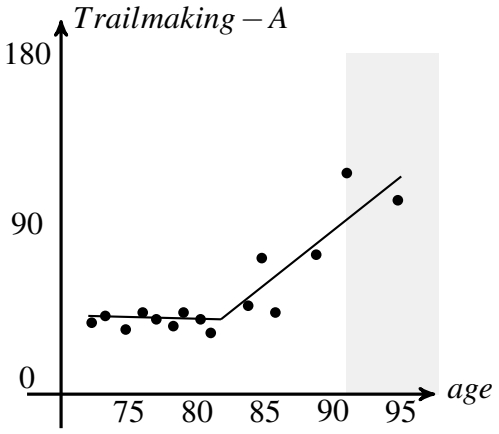

Figure 1.3: Individual Fit

1.5 Long Term Efforts

This thesis lays the foundation for a larger research program aimed at finding the timing of phase change prospectively in Alzheimer's dementia (i.e. a quantitative diagnostic based on longitudinal measurement). As can be seen in **Figure 1.3** and will be evidenced in this thesis there is a substantial period of time between the onset of cognitive decline and clinical diagnosis. By updating our posterior estimates as an individuals data arrives we can characterize the *distribution* of their changepoint in real time. This distribution can then be used for calculating the probability that an individual has transitioned into the disease phase (after the construction of decision rules). This is extremely important because recent research has shown that it may be possible to push back the symptoms of Alzheimer's with early interventions. A few promising interventions include exercise therapy [Erickson et al., 2011], deep brain stimulation [Laxton et al., 2010], and possibly an appli-

cation utilizing stem cells [Blurton-Jones et al., 2009]. For any of these treatments to be of use it would be optimal or necessary to intervene at a point in time before neuron death accumulates to cause massive cognitive and functional problems in the later stages of the disease.

Chapter 2

Data Description

2.1 Introduction

The final analysis consisted of 36 individuals with Alzheimer's Dementia and measurements on at least 9 occasions from archival data in a study of healthy aging and dementia. Positive neuropathology at autopsy or sustained occupancy in non-zero Clinical Dementia Ratings (CDR) were used to determine disease presence (Section 2.2). A summary of the population and the specific measurements used in the analysis are discussed in Section 2.3.

2.2 Determination of Disease Presence

2.2.1 Neuropathology

31% of the individuals used in the analysis had a positive AD neuropathology at postmortem examination. The remaining 69% either did not consent to autopsy or are still alive. All brains were examined with standard protocol[McKeel DW, 1993]. Following fixation in neutral buffered 10% formalin, tissue blocks were taken from 30 brain regions. Sections (6 μm) from paraffin-embedded tissue block were stained with hematoxylin-eosin, and modified silver stains, and immunohistochemical methods[McKeel DW, 1993]. Histologic criteria for AD were based on quantification of

diffuse and neuritic amyloid deposition in five cortical regions with $10mm^2$ microscopic fields in each region as well as National Institute on Aging (NIA) neuropathologic probability estimates of AD. Cases were screened for Lewy bodies with antibodies to alphasynuclein and were also examined for the presence of cortical and subcortical infarcts and hemorrhages to exclude confounding dementia diagnosis.

2.2.2 Clinical Dementia Rating

To determine the presence and severity of dementia the research program staged dementia using the CDR. The CDR evaluates cognitive function in six categories (memory, orientation, judgment and problem solving, performance in community affairs, home and hobbies, and personal care) without reference to psychometric performance or results of previous evaluations. CDR 0 indicates no dementia, and CDR 0.5, 1, 2, or 3 correspond to very mild, mild, moderate, and severe dementia.

In **Table 2.1**(a) the number of adjacent CDR transition of 115 people with at least 2 non-zero CDR and more than 9 measurements are described. This matrix describes how the population transitions through the states (0,0.5,1,2,3) of the CDR. Interpretation is straightforward ($from\ CDR \mapsto to\ CDR$), the number of times that a CDR 0 precedes a CDR 0.5 is 123 [1, 2] and the number of times this relationship is reversed is 82 [2, 1].

(a) Misclassified Data

$\mapsto$	0.0	0.5	1.0	2.0	3.0
0.0	518	123	4	0	1
0.5	82	285	62	3	0
1.0	0	19	102	34	2
2.0	0	0	5	41	7
3.0	0	0	0	0	3

(b) Reclassified Data

$\mapsto$	0.0	0.5	1.0	2.0	3.0
0.0	283	33	2	0	0
0.5	0	56	10	2	0
1.0	0	0	16	4	0
2.0	0	0	0	2	1
3.0	0	0	0	0	1

Table 2.1: CDR Transition Tables

The concern is that in **Table 2.1**(a) individuals transition from a higher to lower severity at a high rate (**without intervention**). Alzheimer's dementia is a progressive dementia thus transitions from higher to lower severity are not realistic. To remedy this these backwards transitions were

treated as misclassification of the true states. For concreteness, an individual in the data has a CDR profile $(0,0,0,0.5,0,0,0,0,0,0.5,0,0,0)$ over a roughly 14 year span. This individuals two CDR 0.5 are treated as misclassifications due to the sparsity of non-zero CDR, i.e. they are considered non-demented and excluded from the AD analysis. This reclassification procedure trimmed the data to 35 individuals. Each person had at least 2 successive non-zero CDR (median = 4) without regression and started the study with a CDR 0. The transition matrix for the reclassified data can be seen in **Table 2.1**(b) . Note that the left diagonal is filled with zeros which indicates forward-only progression. The total number of individuals after this procedure was 35.

To be absolutely clear, there were 115 people in the research program that started as CDR 0, had more than 2 non-zero CDR throughout the duration of the study, and had 9 or more observations. Only 35 of these people fit the AD classification criteria. This procedure independently included all but one of the individuals that showed positive neuropathology at autopsy - the lone individual was never assigned a non-zero CDR(this individual brings N=36)

2.3 Summary of Population and Measurements

Table 2.2 shows the ages of the population at important points in the research program. Of importance is that with the reclassified data there is roughly 11 years of data before the first non-zero CDR and thus the data is well equipped to define changes in pre-clinical AD.

	Mean (SD)
Age at First Measurement	72.68 (7.03)
Age at Non-Zero CDR	83.18 (6.88)
Age at Last Measurement	87.51 (6.97)
Number of Observations	12.33 (4.34)

Table 2.2: Summary Statistics

A 90 minute test battery was administered annually to all participants approximately 2 weeks after clinical evaluation. This battery tests across multiple cognitive domains (i.e. semantic memory, episodic memory, working memory, and visuospatial memory). Tests included in the analysis are Associate Learning, Boston Naming, Logical Memory, Benton Visual Retention Test: Form D-Delayed, Digit Symbol, Trailmaking-A, Block

Design and Word Fluency for S and P. Multiple other tests were included in the battery but not in the analysis due to their scale. Psychometricians were not informed of the results of the clinical evaluation.

Each of the tests are shown relative to clinical diagnosis for the population of 36 AD individuals in **Figure 2.1**. These loess curves should ***not*** be taken as a literal description of the data. Loess estimation is extremely sensitive to outliers, the curves don't have a known functional form, and the data is aligned at CDR - which is suboptimal. Furthermore, loess curves are not appropriate for effectively modeling individual differences. Alignment at CDR and loess were used as an approximation to present the nonlinear nature of the data in a compact manner.

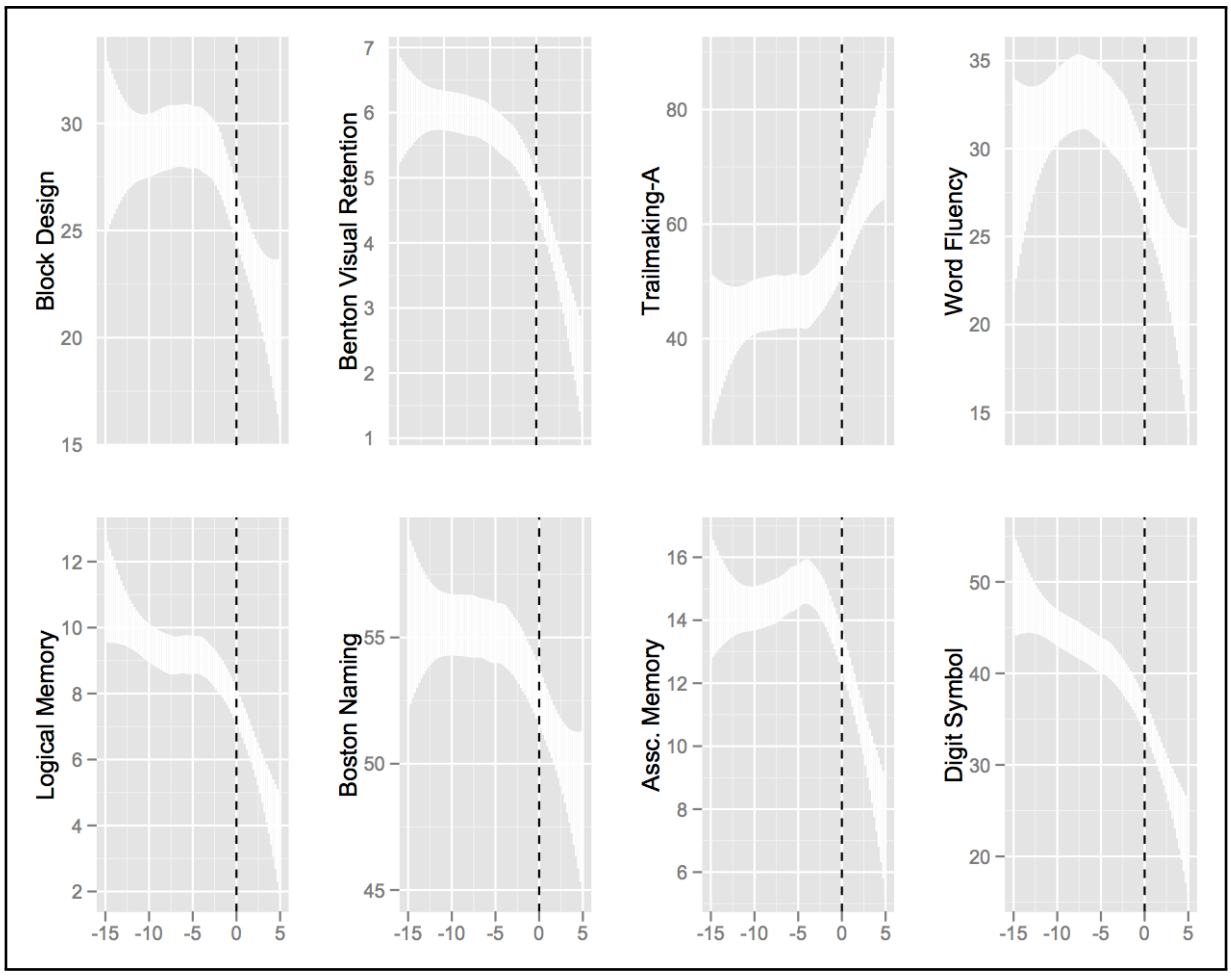

Figure 2.1: with Time Relative to Clinical Diagnosis (Dashed Line)

Chapter 3

Methodology

3.1 Model Specification

It is straightforward to extend Equation 1.1 to incorporate hierarchical dependencies. It first involves adding random effects to the model to account for multiple individuals. Then the hierarchical structure is incorporated by modeling the relationships between the random effects parameters. In Bayesian statistics this is accomplished by considering prior distributions of the random effects parameters themselves to be generated from shared parameters known as hyperparameters(also referred to as population parameters). This process is shown for two classes of multiphase models; an abrupt change process and a smoothed change process.

3.1.1 Abrupt Hierarchical Model

3.1.1.1 Random Effects

To begin we will respecify Equation 1.1 to account for i individuals. Let Y_{ij} be the response vector of outcomes of the j^{th} measurement time of the i^{th} individual and t_{ij} be a corresponding vector of observation times. Assume that Y_{ij} are distributed normally with expectation μ_{ij} and precision τ, $Y_{ij} \sim N(\mu_{ij}, \tau)$. Our equation becomes

$$\mu_{ij} = \begin{cases} \alpha_{i0} + \alpha_{i1} t_{ij} & t_{ij} < \delta_i \\ \alpha_{i0} + \alpha_{i1}\delta_i + \alpha_{i2}(t_{ij} - \delta_i) & t_{ij} \geq \delta_i \end{cases} \tag{3.1}$$

The parameters in this model can be interpreted in the same manner as Equation 1.1 with the exception that there are i individuals. That is, for the ith individual, α_{i0} is the level of Y at the beginning of the the phase before the changepoint δ_i, with α_{i1} and α_{i2} being the first and second phase slopes, respectively.

3.1.1.2 Dependencies

The DAG for the model is illustrated in **Figure 3.1**. It shows the hierarchical dependencies of individual i's parameters to the entire AD population. The point is that an individuals vector of data are generated by a normal distribution determined (dotted lines) by parameters $\alpha_{i0,1,2}$, and δ_i (line 1 in Equation 3.2). These individual level parameters are in turn generated from higher level normal or multivariate normal distributions with hyperparameters $\mu_{\alpha_{0,1}}$, $\Omega_{\alpha_{0,1}}$, μ_{α_2}, τ_{α_2}, μ_δ and τ_δ. (lines 2-4 in Equation 3.2). This indicates that each individuals parameters come from common population distributions. Given that we want to estimate the parameters of the AD population, these higher level hyperparameters are given hyperpriors (lines 5-9 in Equation 3.2). Line 10 is simply the prior assigned to τ, the precision of Y_{ij}. The full specification is given by

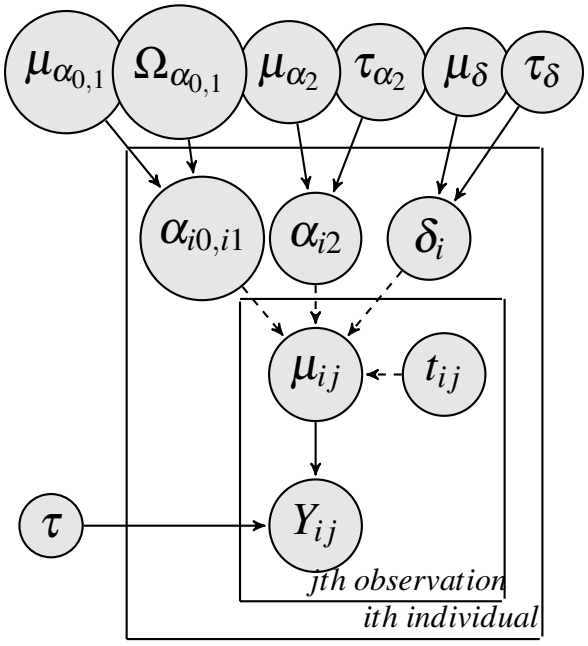

Figure 3.1: Directed Acyclic Graph: Abrupt Model

$$
\begin{aligned}
Y_{ij}|\alpha_{i0,1,2},\delta_i,t_{ij},\tau &\sim N((\alpha_{i0,1,2},\delta_i,t_{ij}),\tau) \\
\alpha_{i0,1}|\mu_{\alpha_{0,1}},\Omega &\sim MVN((\mu_{\alpha_{0,1}}),\Omega_{\alpha_{0,1}}) \\
\alpha_{i2}|\mu_{\alpha_2},\sigma_{\alpha_2} &\sim N(\mu_{\alpha_2},\tau_{\alpha_2}) \\
\delta_i|\mu_\delta,\sigma_\delta &\sim N(\mu_\delta,\tau_\delta) \\
\Omega_{\alpha_{0,1}} &\sim Wishart\left(\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix},2\right) \\
\mu_{\alpha_{0,1,2}} &\sim N(0,0.001) \\
\tau_{\alpha_2} &\sim G(0.001,0.001) \\
\mu_\delta &\sim N(0,0.001) \\
\tau_\delta &\sim G(0.001,0.001) \\
\tau &\sim G(0.001,0.001)
\end{aligned}
\tag{3.2}
$$

3.1.1.3 Discussion of Prior Choices

α_{i0} and α_{i1} were modeled as being generated from a multivariate normal distribution with hyper-parameters μ_{α_0} and μ_{α_1} and Ω. The multivariate normal was used because it is common for these two parameters to be correlated and thus to complicate convergence (it makes large jumps difficult in an MCMC chain). Ω was given a conjugate prior to the multivariate normal, namely the inverse Wishart(R,ρ). To represent vague prior knowledge the degrees of freedom were as small as possible $\rho = 2$. The scale matrix $R = \left(\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right)$ is a guess at the magnitude of the covariance matrix of $\alpha_{0,1}$. Lindley [Lindley, 1970] showed that the choice of R had little effect on the posterior estimates. G.(shape ,scale) represents the gamma distribution and N represents the normal with N(mean, precision). The use of precision is just an artifact of the Gibbs sampler. These are commonly used diffuse or *non-informative* priors for regression and variance parameters.

3.1.2 A Smoothed Model

A second type of multiphase model put forth by Bacon and Watts [Bacon & Watts, 1971] was estimated for purposes of model comparison. This model used a hyperbolic tangent function to smooth the transition between the two phases. It is often a more appropriate model if an abrupt transition, as in Equation 3.1, does not take place.The DAG is shown in **Figure 3.2** - note that the difference between the two models is the new parameter γ. The smoothed model is given by

$$\mu_{ij} = \alpha_{i0} + \alpha_{i1}(t_{ij} - \delta_i) + \alpha_{i2}(t_{ij} - \delta_i)tanh((t_{ij} - \delta_i)/\gamma) \tag{3.3}$$

Following the specification that was given in [van den Hout et al., 2011] γ was assigned a uniform prior with bounds 0 and 10. The important point to note is that as γ approaches zero a more abrupt transition occurs and as it grows larger a smoother transition occurs. This will become important in the comparison of the models in the next chapter. The rest of the parameters were assigned the same priors as in the abrupt model. A difficulty with this model is in the interpretation of α_1 and α_2, in that these are not the same as in Equation 3.1 but are tied into the estimate of γ.

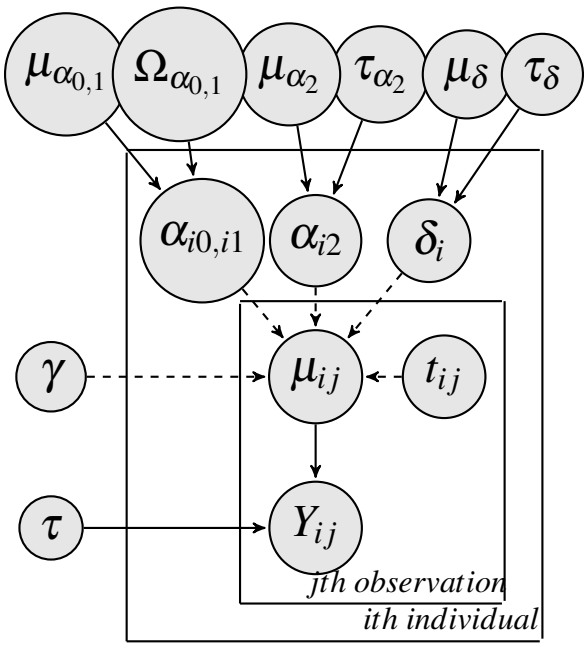

Figure 3.2: Directed Acyclic Graph: Smooth Model

3.2 Alignment Procedure

There is a caveat to the model specification as described thus far. The central idea behind the modeling in this analysis is to define patterns of cognitive decline in AD. To this end a procedure was defined to anchor each of the individual's 8 outcomes (Section 2.3) to their changepoint estimate

of Logical Memory (LM), estimated at the scale of age. This redefines the time of each of the measurements as a difference, $\dot{t}_{ij} = t_{ij}$ - $\delta_{i(LM)}$. Logical Memory was chosen because it is a well understood outcome in AD. The flexibility of the Bayesian framework allows the estimation of all the parameters in concert and thus to use the distribution $\delta_{i(LM)}$ in the calculation of $\dot{t}_{ij}$. This specification means that $\dot{t}_{ij}$ is dependent on the precision of the $\delta_{i(LM)}$ and thus carries the uncertainty associated with it.

The caveat being that the prior specification in Equation 3.2 should be made more appropriate for the initial estimation of age at $\delta_{i(LM)}$. To do this we placed a normal distribution on the location hyperparameter $N(75, 0.001)$ and the usual gamma on the precision. The aforementioned priors in Equation 3.2 stay the same for the other 7 outcomes, and then the re-estimation of Logical Memory on $\dot{t}_{ij}$. As a check, for the procedure to be reasonable the re-estimation of Logical Memory should be centered very close to 0 (actual estimate of the mean is 0.003). To be clear, all of the outcomes were modeled simultaneously, using each individuals changepoint estimate of LM over age to define a new time scale which aligns everybody at their LM changepoint (the exact same changepoint/time scale for each outcome).

3.3 Bayesian Estimation in Practice

The estimation of our hierarchical models follows the same scheme as was shown in the introductory chapter - simulation of posteriors using Gibbs sampling. JAGS [Plummer & Plummer, 2003] was the Gibbs sampler used to simulate posterior densities. JAGS works by constructing a DAG, similar to the graphs in this chapter, that represents the parameters and conditional independencies of a model. Because of its factorization (see: [Lauritzen & Spiegelhalter, 1990]) a DAG permits a complex full joint distribution to be specified in terms of its local components(conditional distributions). It is then possible to set up an MCMC algorithm to sample marginal distributions using these conditional distributions. Code is given in the appendix. There are two practical issues that need to be addressed; the first is how to determine when an MCMC chain has converged to the

target distribution, and the second is model selection with Deviance Information Criterion (DIC).

3.3.1 Convergence of Simulation

The use of multiple MCMC chains (2-5) with lengths between 10,000 and 50,000 were used to summarize the marginal posteriors of interest. The Gelman-Rubin convergence statistic was used to detect convergence of parameters [Gelman & Rubin, 1992]. The Gelman-Rubin statistic quantifies the difference in the variance within a chain (W) and between all chains (B) to detect convergence. It is based on the basic analysis of variance. When the between chain variance is no larger than the within chain variance then it has *probably* converged on a solution. As a simulation converges the statistic $\hat{r} = B/W$ approaches 1.0. A general rule of thumb set forth in [Gelman et al., 2003] is that $\hat{r}$ be less than 1.2 to be considered a stable estimate. This diagnostic was monitored for all parameters and kept below the conservative value of 1.05 - well below the rule of thumb. That is, if an estimate was above 1.05 more chains or iterations were run. Trace plots were also examined to ensure proper mixing of the chains as to not take a simulation that gave a false positive on the Gelman-Rubin diagnostic.

3.3.2 Deviance Information Criterion

Ultimately, the parameter estimates shown in the next chapter come from simultaneous estimation of all the outcomes. However, each outcome was modeled separately to get unique Deviance Information Criterion(DIC) [Spiegelhalter et al., 2002] values because some of the outcomes might follow a smoothed transition and some an abrupt. DIC is a hierarchical generalization of Akaike's Information Criterion(AIC). It can be used to test the fit of different prior specifications and non-nested models. Similar to AIC, a model with lower DIC has a higher chance of predicting a replicate data set. A difference more than 5 is generally considered to be substantial and more then 10 should definitely rule out the model with the higher DIC.

Chapter 4

Model Selection and Estimates

4.1 Model Selection

Estimates of DIC are given in **Table 4.1**. It can be seen that the abrupt model outperforms the smooth model in all instances (lower DIC being desirable). The difference in DIC between the abrupt and smoothed models can be explained in two interrelated parts. First, the estimates of the population changepoint are similar between the smooth and the abrupt models and the γ estimates of the smoothed model are sufficiently close to zero to signal an abrupt change process. Given DIC penalizes for the effective number of parameters, the more simplistic abrupt model is favored. Second, and related to the first, is that, while the parameter estimates are similar, they are not exactly the same and simply do not replicate the data as well (i.e. DIC difference is not only a function of effective parameters). The general tendency is for the smoothed model to estimate μ_δ to be a little further

	DIC	
	Abrupt	Smoothed
Block Design	4306	4314
BVRT-Delayed	3163	3168
Trailmaking A	5456	5463
Word Fluency	4415	4421
Logical Memory	3439	3452
Assc. Memory	3712	3720
Boston Naming	3671	3682
Digit Symbol	4510	4519

Table 4.1: DIC for Abrupt and Smooth Models

away from Logical Memory and for μ_{α_2} to be a little less graded than in the abrupt model.

While the smoothed model is a more general model (allowing more types of change) there is no reason to favor it if the process being studied is, in fact, an abrupt change process. It appears that cognitive decline in AD is one such process. The parameter estimates in the rest of this section are devoted to the abrupt model.

4.2 Population Estimates

4.2.1 Estimates of the Changepoint

Estimates of δ for each of the measurements are included in **Table 4.2**. The mean and standard deviation of the population changepoint (far right column) are calculated from the hyperparameters μ_δ and μ_τ. The mean is simply the peak of μ_δ and the standard deviation is calculated while sampling μ_τ in JAGS as $\sigma = \frac{1}{\sqrt{\tau}}$. That is to say the most probable distribution to describe the population δ is a normal distribution with mean being the peak of μ_δ and precision as the peak of μ_τ. For example, the most probable changepoint distribution for Block Design is a normal distribution located 3.97 years before Logical Memory and with a standard deviation of .63.

	95% Highest Density Interval		
	μ_δ	μ_τ	$\delta(\sigma)$
Block Design	(-5.81, -2.19)	(0.16, 6.88)	-3.97 (0.63)
BVRT-Delayed	(-5.20, -1.99)	(0.12, 5.69)	-3.55 (0.71)
Trailmaking A	(-3.53, -1.51)	(2.99, 17.79)	-2.76 (0.30)
Word Fluency	(-1.56, 0.97)	(0.04, 5.73)	-0.42 (0.84)
Logical Memory	(-1.01, 1.12)	(0.66, 8.97)	-0.00 (0.48)
Assc. Memory	(-0.92, 2.23)	(0.16, 6.53)	0.72 (0.67)
Boston Naming	(0.46, 2.51)	(0.40, 7.88)	1.52 (0.61)
Digit Symbol	(0.49, 2.50)	(0.04, 4.68)	1.45 (0.96)

Table 4.2: 95% Highest Density Interval for μ_δ and μ_τ with Mean and Standard Deviation Calculated.

The simulated distribution for μ_δ can be seen in **Figure 4.1**, grouped according to a previous

confirmatory factor analysis of the data [Johnson et al., 2008]. This figure describes the sampled posterior mean (with the ticks at the bottom being draws from the posterior) of the changepoint estimated pattern of cognitive decline. The three memory types shown in our previous research were visuospatial, verbal, and working memory.

4.2.2 Estimates of the Regression Parameters

The population estimates of α_0, α_1, and α_2 are given in **Table 4.3**. These were calculated in the same manner as were the δ estimates in the preceding section. All of these parameters behave as would be expected. With the exception of Trailmaking-A and Word Fluency all of the first phase slopes are slightly declining - relative to the second phase. Word Fluency can be considered to stay flat and Trailmaking-A is on a scale in the opposite direction of the rest of the measurements. All of these demarcate a slightly worsening or non-changing process consistent with healthy normal aging (detailed later and shown in **Appendix B**) and then a change in phase that marks the onset of the disease process as measured by each test. It is important to note that none of the α_1 and their respective α_2 distributions contain each other. This indicates a meaningful switch from the first to the second phase.

	Parameter		
	α_0	α_1	α_2
Block Design	28.60 (1.19)	-0.15 (0.09)	-2.37 (0.46)
BVRT-Delayed	5.49 (0.25)	-0.05 (0.02)	-0.37 (0.11)
Trailmaking A	51.83 (1.36)	0.99 (0.17)	10.95 (1.33)
Word Fluency	31.57 (1.70)	0.01 (0.13)	-2.12 (0.51)
Logical Memory	8.52 (0.52)	-0.04 (0.03)	-0.89 (0.18)
Assc. Memory	14.39 (0.80)	-0.05 (0.04)	-1.90 (0.70)
Boston Naming	52.66 (0.74)	-0.19 (0.05)	-8.73 (1.15)
Digit Symbol	37.47 (1.66)	-0.69 (0.12)	-5.85 (1.06))

Table 4.3: Population Estimates of Intercept, Phase 1, and Phase 2 Slopes

4.3 Individual Estimates of the Changepoint

Previous studies of phase change in Alzheimer's disease have adopted the method of alignment of individuals onto their first non-zero CDR. This analysis did not use this method because of variability in the assignment of CDR. This will possibly lead to incorrect inferences in the models parameters and hyperparameters and give false patterns of the changepoints or distort the timing relationships between them. That said, it is still possible to calculate the difference between any of the measurements and the individuals first non-zero CDR. A posterior distribution of the time between an individuals changepoint on Block Design and the first non-zero CDR was sampled - the mean of this distribution is -6.05 (1.32) years prior to CDR. This was calculated by taking the time between the Logical Memory changepoint and their CDR and then adding the time between Logical Memory and Block Design.

The calculated mean of the Logical Memory changepoint on the age scale was 80.89 years of age. The average age of first non-zero CDR was 83.18 years - a difference of 2.29 years(roughly what should be expected, given the mean δ of Block Design is -3.97 years relative to Logical Memory). A plot of the individual changepoint estimates of Logical Memory on the age scale are shown in Figure 4.2. As evidenced in this plot, for all but the second and fifteenth individual, the CDR comes after or within the bounds of the estimated changepoint (not counting the individual who had no non-zero CDR but showed positive neuropathology).

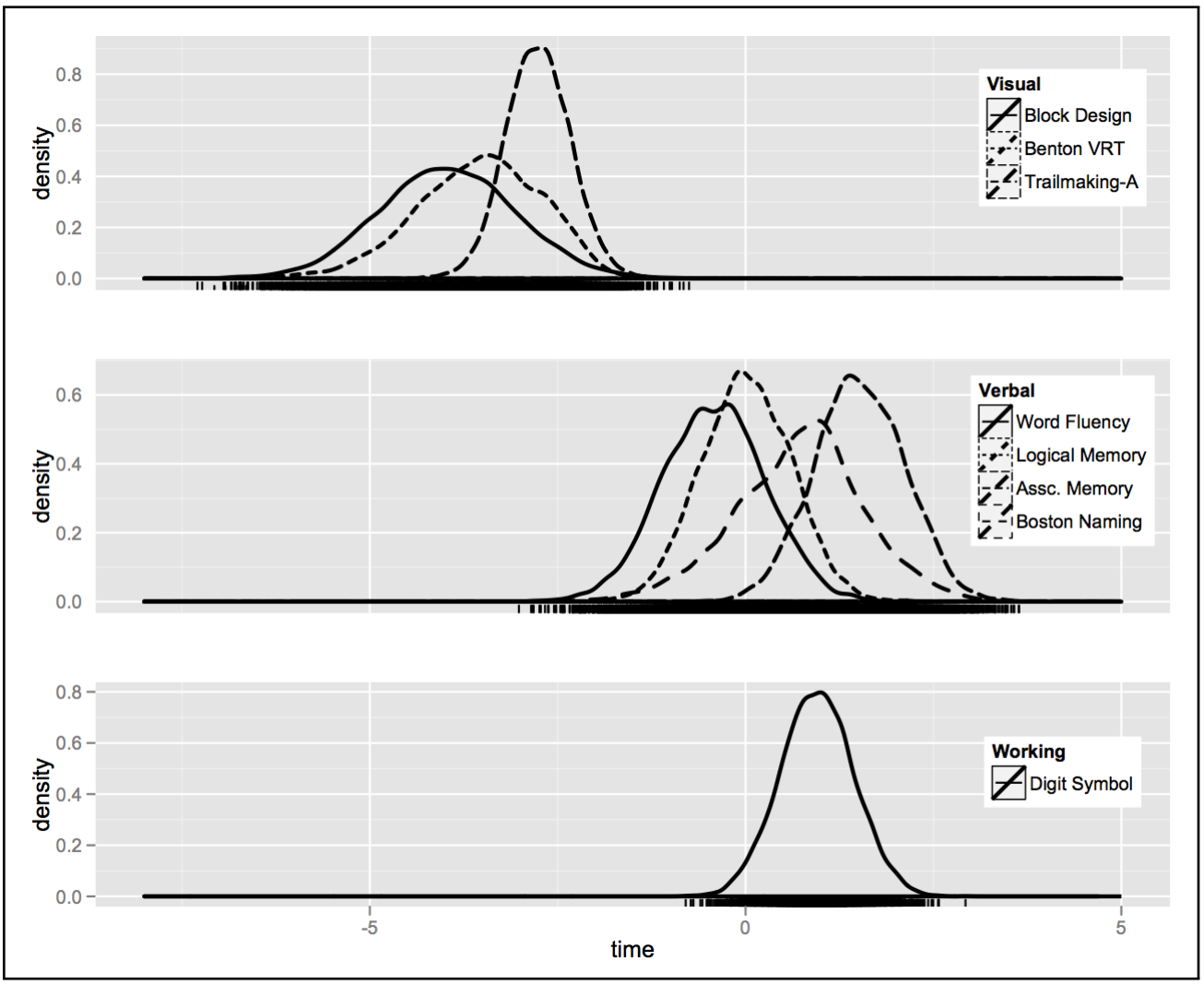

Figure 4.1: Simulated Distributions of μ_δ Ordered by Memory Type.

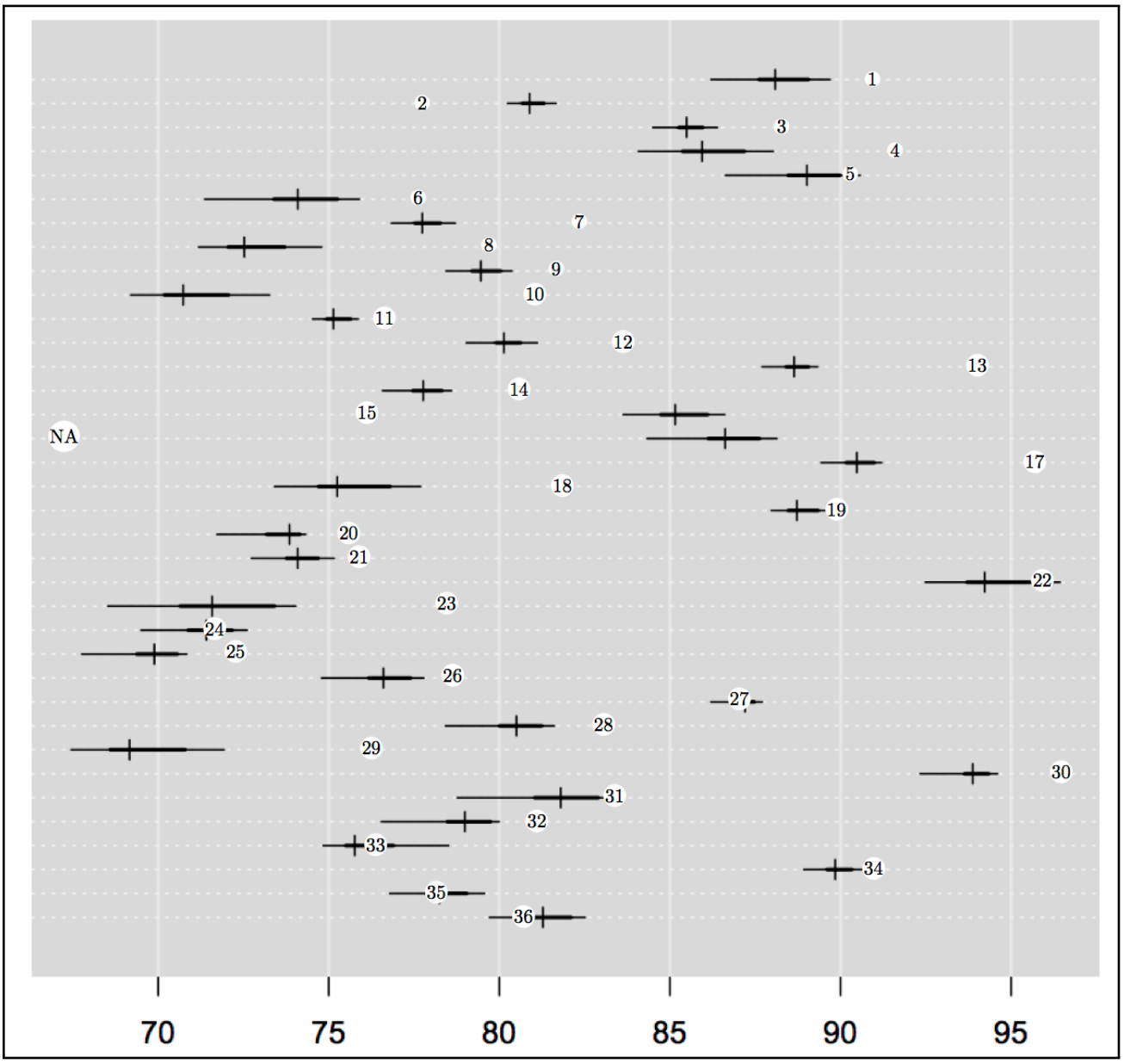

Figure 4.2: 68% and 95% Credible Intervals for Individual δ Estimates of Logical Memory. Each Individuals Age of First Non-Zero CDR is marked with a white circle. NA is for the individual without a non-zero CDR score.

Chapter 5

Discussion

5.1 Overview

In this thesis we demonstrated the use of Bayesian estimation for two classes of multiphase models - an abrupt change process and a smoothed process. Due to the similarity in the parameter estimates (and differences in DIC) between the abrupt and smoothed model the more sparsely paramaterized abrupt model was chosen for the full analysis. That means, in either instantiation of a phase change model the interpretation that cognitive decline in AD is an abrupt process is supported.

If we calculate through Block Design we see that *by the time* an individual is diagnosed the level of the outcome has dropped by 14.34 points on average. This drop in level comes after the first phase that starts at a level of 28.60(1.19) which is similar to the intercept estimate of 28.36(1.49) for a linear hierarchical model of healthy older adults(shown in **Appendix B**). Furthermore the corresponding slopes are within each others bounds. Given this, we argue that *where* this decline starts can be considered a significant *milestone* of AD - the transition time from healthy to diseased cognition. By comparing **Appendix B** with the mean estimates of α_0 and α_1 in **Table 4.3** we can see that this holds across all the measurements. The first phase in AD simply cannot be differentiated from the healthy older adults.

The mean pattern of cognitive decline in AD can be seen in Figure 4.1. By using a method

that anchored the changepoints relative to Logical Memory we have defined a pattern that holds irrespective of age and clinical status (the method used in most analysis). That is, at any point in time that someone develops (late onset) AD the most probable pattern that their cognitive decline will follow is the pattern that we have produced. It shows that there are definite cognitive 'clusters' of decline. The first cluster is the visuospatial domain that begins on average 3.97(.63) years before Logical Memory. The next cluster is the verbal memory domain, all of these verbal memory tests begin there phase change at roughly the same time as Logical Memory. Working memory abilities, as indicated by Digit Symbol, appear to come at the tail end of verbal memory. More evidence by way of more outcomes is needed to make strong remarks about the working memory domain.

5.2 Relationship to Past Research

This pattern is mostly in agreeance with a previous publication [Johnson et al., 2009] that used a fixed effects method to estimate the most likely location of a pre-specified changepoint by profiling the likelihood. The profile-likelihood method reported that visuospatial abilities started to decline 3 years before CDR, and both verbal and working memory abilities declined 1 year before CDR. It is a similar pattern in which visuospatial memory is the first to decline. But, the current method, as was shown in the last chapter puts the time of change in Block design 6.05(1.32) years before CDR. There are numerous reasons that could explain these differences; the criteria used to define a non-diluted sample, the method of alignment, or the ability to account for individual differences are the most plausible.

5.3 Future Efforts

As was detailed in this thesis visuospatial tests begin to decline numerous years prior to clinical diagnosis. This is a significant amount of time between, what has been argued is, the cognitive onset of the disease and the formal diagnosis by a clinician. In the prostate specific antigen literature [Slate & Turnbull, 2000] similar models have been used to quantify the changepoint prospectively

in prostate cancer. The idea being to develop an algorithm that monitors the posterior distribution of the changepoint as an individuals data is captured in a manner that classifies AD early and correctly. The models could prove useful because they not only take into account where an individual is located at the current moment, but also where they were at times prior. There is another method that could possibly accomplish this same goal - the Viterbi algorithm in hidden markov models. Future work will be aimed at developing these multiphase and hidden markov models for prospective detection of AD. From this thesis it is understood that an intelligent starting point would be visuospatial memory - though the methods are not constrained to a single test (in fact, detection should be better by using multiple tests).

References

[Bacon & Watts, 1971] Bacon, D. & Watts, D. (1971). Estimating the transition between two intersecting straight lines. *Biometrika*, 58(3).

[Blurton-Jones et al., 2009] Blurton-Jones, M., Kitazawa, M., Martinez-Coria, H., Castello, N., Müller, F.-J., Loring, J., Yamasaki, T., Poon, W., Green, K., & LaFerla, F. (2009). Neural stem cells improve cognition via bdnf in a transgenic model of alzheimer disease. *Proceedings of the National Academy of Sciences of the United States of America*, 106(32), 13594–13599.

[Erickson et al., 2011] Erickson, K., Voss, M., Prakash, R., Basak, C., Szabo, A., Chaddock, L., Kim, J., Heo, S., Alves, H., White, S., Wojcicki, T., Mailey, E., Vieira, V., Martin, S., Pence, B., Woods, J., McAuley, E., & Kramer, A. (2011). Exercise training increases size of hippocampus and improves memory. *Proceedings of the National Academy of Sciences of the United States of America*, 108(7), 3017–3022.

[Gelman et al., 2003] Gelman, A., Carlin, J., Stern, H., & Rubin, D. (2003). *Bayesian data analysis*. Chapman and Hall/CRC, 2 edition.

[Gelman & Rubin, 1992] Gelman, A. & Rubin, D. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4), 457–472.

[Geman & Geman, 1984] Geman, S. & Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, PAMI-6(6), 721–741.

[Johnson et al., 2009] Johnson, D., Storandt, M., Morris, J., & Galvin, J. (2009). Longitudinal study of the transition from healthy aging to alzheimer disease. *Arch Neurol*, 66(10), 1254–1259.

[Johnson et al., 2008] Johnson, D., Storandt, M., Morris, J., Langford, Z., & Galvin, J. (2008). Cognitive profiles in dementia: Alzheimer disease versus nondemented healthy brain aging. *Neurology*, 71(22), 1783–1789.

[Kiuchi et al., 1995] Kiuchi, A., Hartigan, J., Holford, T., Rubinstein, P., & Stevens, C. (1995). Change points in the series of t4 counts prior to aids. *Biometrics*, 51(1), 236–248.

[Klein JP, 1984] Klein JP, Klotz JH, G. M. (1984). A biological marker model for predicting disease transitions. *Biometrics*.

[Lauritzen & Spiegelhalter, 1990] Lauritzen, S. & Spiegelhalter, D. (1990). Local computations with probabilities on graphical structures and their application to expert systems. *Readings in Uncertain Reasoning*, (pp. 415–448).

[Laxton et al., 2010] Laxton, A. W., Tang-Wai, D. F., McAndrews, M. P., Zumsteg, D., Wennberg, R., Keren, R., Wherrett, J., Naglie, G., Hamani, C., Smith, G. S., & Lozano, A. M. (2010). A phase i trial of deep brain stimulation of memory circuits in alzheimer's disease. *Annals of Neurology*, 68(4), 521–534.

[Lindley, 1970] Lindley, D. V. (1970). *The estimation of many parameters*. Holt, Rinehart and Winston, Toronto.

[McKeel DW, 1993] McKeel DW, Ball MJ, P. J. (1993). Interlaboratory histopathologic assessment of alzheimer neuropathology: different methodologies yield comparable diagnostic results. *Alzheimer Dis Assoc Disord*, 7(136–151).

[Plummer & Plummer, 2003] Plummer, M. & Plummer, M. (2003). Jags: A program for analysis of bayesian graphical models using gibbs sampling.

[Slate & Turnbull, 2000] Slate, E. H. & Turnbull, B. W. (2000). Statistical models for longitudinal biomarkers of disease onset. *Statistics in Medicine*.

[Spiegelhalter et al., 2002] Spiegelhalter, D., Best, N., Carlin, B., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4), 583–639.

[van den Hout et al., 2011] van den Hout, A., Muniz-Terrera, G., & Matthews, F. (2011). Smooth random change point models. *Statistics in medicine*, 30(6), 599–610.

Appendix A

JAGS Code

This is code for a simple univariate linear-linear model. Multiple outcomes can be easily aligned by creating an alignment time scale as discussed in Chapter 3.

```
 model {
for (j in 1:N){
for (k in ind[j]:(ind[j+1]-1)){
      Y[k]~dnorm(mu[k],tau)
      mu[k]<-(b[j,1]+b[j,2]*t[k])*(1-step(t[k]-b[j,4])) //
+ (b[j,1]+b[j,2]*b[j,4]+b[j,3]*(t[k]-b[j,4]))*step(t[k]-b[j,4])
#step() returns 1 if argument is >= 0, 0 otherwise
          }}
    for(j in 1:N){
     b[j,1:2]~dmnorm(mu.b[1:2],T.b[1:2,1:2])
     b[j,3]~dnorm(mu.b[3],tau.b[3])
     b[j,4]~dnorm(mu.b[4],tau.b[4])
      }

# variance parameters
```

```
 tau~dgamma(.001,.001)
 tau.b[3]~dgamma(.001,.001)
 tau.b[4]~dgamma(.001,.001)
 Sigma1[1:2,1:2] <- inverse(T.b[,])
 sigma[1]<-sqrt(Sigma1[1,1])
 sigma[2]<-sqrt(Sigma1[2,2])
 sigma[3]<-1/sqrt(tau.b[3])
 sigma[4]<-1/sqrt(tau.b[4])

# location parameters
 for(j in 1:4){ mu.b[j]~dnorm(0,.001)  }
 # or mu.b[4]~dnorm(75,0.001)

 # Wishart
 R[1,1] <- 1; R[1,2] <- 0;
 R[2,1] <- 0; R[2,2] <- 1;
 T.b[1:2,1:2] ~ dwish(R[,], 2)
 }
```

Appendix B

Healthy Older Adults

Below is a table for the intercept and slope of a hierarchical linear model for 23 individuals with non-positive neuropathology at autopsy. These are the population estimates, similar to the rest of the models in the thesis.

	Parameter	
	α_0	α_1
Block Design	28.36 (1.49)	-0.21 (0.12)
BVRT-Delayed	5.12 (0.62)	-0.07 (0.09)
Trailmaking A	54.30 (1.78)	1.23 (0.27)
Word Fluency	32.82 (2.10)	-0.10 (0.22)
Logical Memory	9.01 (0.88)	-0.00 (0.16)
Assc. Memory	13.58 (1.01)	-0.06 (0.17)
Boston Naming	53.92 (0.85)	-0.17 (0.22)
Digit Symbol	39.11 (2.03)	-0.61 (0.33)